Solution for "ICDM 2022 : Risk Commodities Detection on Large-Scale E-commerce Graphs"

Xiaocheng Yang, Mingyu Yan

ICT-GimLab, Institute for Computing Technology, Chinese Academy of Sciences, Beijing, China
yangxiaocheng@ict.ac.cn, yanmingyu@ict.ac.cn

Technical Report

Model

We use the basic SeHGNN (Yang et al. 2022) model for this contest. We tested many variations of the model which have displayed better performance in the validation set, but the basic model over-performs others in the test sets of both session 1 and session 2.

Network structure

For metapath selection, we utilize all metapaths no more than 4 hops (i.e., *i, ib, if, ibi, ifi, ibif, ifae, ifib, ifaf, ifcf, ifdf, ifef*). For the feature projection block, we use a single-layer MLP for each metapath. For the task-specific module, we use a two-layer MLP. The dimension of hidden vectors is 512.

We remove the label propagation and multi-stage training modules of SeHGNN due to the requirements of this contest.

Training configure

The model is optimized with Adam during training. The learning rate is 0.001 and the weight decay is 0. We use the early-stop mechanism that the training would stop if the best validation average precision (val_ap) does not increase in 50 epochs. We choose the network parameters that achieves the highest val_ap during training.

Trick

We find around 15% positive nodes of the validation set cannot be correctly classified for almost all models. So we re-split the dataset by adding these nodes to the training set. This trick brings significant increase (around 1%) on model effects.

	Average precision
SeHGNN	90.15±0.33
SeHGNN+re-split	91.09±0.20

After applying this trick, the val_ap always achieves nearly 100% so that we can learn nothing from ablation experiments.

References

Yang, X.; Yan, M.; Pan, S.; Ye, X.; and Fan, D. 2022. Simple and Efficient Heterogeneous Graph Neural Network. *arXiv preprint arXiv:2207.02547*.